Analysing online reviews to investigate customer behaviour in the sharing economy: The case of Airbnb

Journal:	*Information Technology & People*
Manuscript ID	ITP-10-2018-0475.R2
Manuscript Type:	Article
Keywords:	Text analysis < Qualitative method < Method, Consumer behaviour < Theoretical concept, Social media < Technology, Business Intelligence < Management practices < Practice

Analysing online reviews to investigate customer behaviour in the sharing economy: The case of Airbnb

Abstract

Purpose – This paper aims to investigate attributes that influence Airbnb customer experience by analysing online reviews from users staying in London. It presents a text mining approach to identify a set of broad themes from the textual reviews. It aims to highlight the customers' changing perception of good quality of accommodations.

Design/methodology/approach – This paper analyses 169,666 reviews posted by Airbnb users who stayed in London from 2011 to 2015. Hierarchical clustering algorithms are used to group similar words into clusters based on their co-occurrence. Longitudinal analysis and seasonal analysis are conducted for a more coherent understanding of the Airbnb customer behaviour.

Findings – This paper provides empirical insights about how Airbnb users' mind-set of good quality of accommodations changes over a 5-year timespan and in different seasons. While there are common attributes considered important throughout the years, exclusive attributes are discovered in particular years and seasons.

Research limitations/implications – This paper is confined to Airbnb experiences in London. Researchers are encouraged to apply the proposed methodology to investigate Airbnb experiences in other cities and detect any change in customer perception of quality stay.

Practical implications – This paper offers implications for the prioritisation of customer concerns to design and improve services offerings and for alignment of services with customer expectations in the sharing economy.

Originality/value – This paper fulfils an identified need to examine the change in customer expectation across the timespan and seasons in the case of Airbnb. It also contributes by illustrating how big data can be used to uncover key attributes that facilitate the engagement with the sharing economy.

Keywords: Airbnb, sharing economy, text mining, consumer behaviour, online review

Article Type: Research paper

1. Introduction

The ongoing digitalisation has significantly spurred new business models since the recent decades. Unlike traditional business models that are based on ownership, the sharing economy is a business model "based on sharing underutilized assets from spaces to skills to stuff for monetary or non-monetary benefits" (Botsman, 2013). The principle is to leverage information technology to empower each party with information that enables distribution, sharing and reuse of excess capacity in goods and services (Heinrichs, 2013). The rise of the sharing economy can be attributed to the paradigm change in consumer behaviour (Puschmann & Alt, 2016).

Consumers are moving away from owning goods to temporarily using goods and making goods available to strangers online (Bardhi & Eckhardt, 2012; Belk, 2013).

In the sharing economy, most individuals have little prior experience with one another before they agree to share. Therefore, many peer-to-peer internet platforms allow users to post public reviews about one another so that other users can be able to make better decisions based on the reviews. While prior studies have investigated how ownership and possession practices shifted towards non-ownership consumption, to our knowledge there is less discussion on how organisations can create value by discovering knowledge from their public reviews, despite the fact that customer-generated big data from digital platforms are important for decision making (Troisi et al., 2018). Motivated by this, this study makes an attempt to mining public reviews and exploiting the discovered knowledge to provide meaningful implications for practitioners.

Due to its dominant position in the sharing economy, Airbnb was chosen in this study. It started in 2008 as a traditional accommodation booking website that combines economic benefits for hosts and guests. To express interest in any of the listings, the guest sends a request or message to the host, followed by direct communication between the host and the guest before a reservation is made. After each stay, both the guest and the host are encouraged to post a review about their experience on the publicly available host and guest pages, respectively (Zervas et al., 2017). Such a review feature is the key trust mechanism of Airbnb (Guttentag, 2015). It allows both hosts and guests to learn more about one another before agreeing to a transaction and also creates incentive for both parties to conduct themselves in an acceptable manner (Jøsang et al., 2007).

Prior studies have confirmed that online reviews influence purchasing decisions, including accommodation booking intention (Sparks & Browning, 2011; Lee & Shin, 2014; Gavilan et al., 2018). As such, analysing reviews is important to discover hidden knowledge that can explain consumer behaviour (Basili et al., 2017). In the context of Airbnb, guests review others' comments and accommodation evaluations before their trips. Their expectation of the stay is affected by the earlier shared experiences they found in the reviews. As a result, it is not uncommon that information discrepancy occurs during their stay. Information discrepancy comes from the gap between their expectation and actual experiences (Bae et al., 2017). Most guests reflect their stay experience and information discrepancy in the reviews that are posted publicly after the completion of their stay. Therefore, most of the reviews contain valuable information about guests' mind-set of good quality of accommodations. By analysing non-textual data, such as average ratings given by customers, one can only get an overview on the size of the gap between customers' expectation and actual experiences. Using the textual data, however, can allow one to gain a deeper understanding of the customers' experiences, including how the experiences differ from the expectation in various aspects. We believe that textual reviews, rather than statistical data, contain personal narratives of experiences made with the Airbnb services. Hence, Airbnb's textual reviews are chosen to be the unit of analysis in this study for uncovering important Airbnb's service attributes that are expected by customers. Our dataset contains 169,666 reviews posted by guests who used Airbnb services in London from 2011 to 2015. Text mining is conducted to identify a set of broad themes from the textual reviews. We believe that change in themes in the reviews may indicate guests' changing mind-set of good quality of accommodations. Firstly, text mining is applied to each year's dataset. Themes are clusters of similar words based on co-occurrence. The purpose is to see how themes in the reviews change over the 5-year timespan. Secondly, each year's dataset is split into multiple sub-datasets based on seasons. The motivation is to investigate if guests' expectation changes when they visit London in different climates.

The rest of the paper is organized as follows. Section 2 presents literature related to this study. Section 3 describes the methodology of this study. Section 4 presents and discusses the results of our longitudinal analysis and seasonal analysis. Lastly, Section 5 concludes this study and provides future research directions.

2. Literature Review

2.1 Airbnb in the sharing economy

The sharing economy is a fast-growing phenomenon. It was highlighted by Time Magazine as one of the ten ideas transforming the world (Walsh, 2011). It is fundamentally stemmed from the concept of pseudo-sharing that was defined by Belk (2014a) as a "phenomenon whereby commodity exchange and potential exploitation of consumer co-creators present themselves in the guise of sharing". There is a variety of forces motivating people to share, including convenience, cost saving, friendship establishment and environmentally friendliness (Cohen & Kietzmann, 2014; Hamari et al., 2016; Rowe, 2016; Tussyadiah & Pesonen, 2016). In addition, the popularity of social media enables the interaction among people who have the desire to share. New communicative technologies allow everyone to engage in entrepreneurial activities and become a provider of all sorts of products and services at the click of a button (Oskam & Boswijk, 2016). For example, Uber drivers can add themselves to the available supply of drivers with a swipe on an app (Zervas et al., 2017).

Airbnb is typically considered an example of the sharing economy. Founded in San Francisco in 2008 as a peer-to-peer accommodation platform, it offers access to millions of places to stay in more than 191 countries, from apartments and villas to castles, treehouses and bed-and-breakfast (Airbnb 2018a). It connects hosts and guests by sharing part or all of homes as rental properties for short stay (Ju et al., 2018). To use the Airbnb services, one must register an Airbnb account. A guest searches based on destination, travel dates and party size, then the website returns a list of available spaces that can be refined by attributes such as price, amenities and property type (Guttentag & Smith, 2017). When a booking is confirmed, Airbnb charges a service fee to both hosts and guests (Airbnb, 2018b).

On the other hand, positioning Airbnb as part of the sharing economy faced criticism (Gunter, 2018). Sharing is about more efficient use of underutilized assets, but it is questionable that housing is an underutilized asset. Oskam and Boswijk (2016) viewed renting out the space to guests as a substitute use rather than additional use. Moreover, due to the monetary nature of Airbnb, *sharing* is a misnomer as it infers altruism. Billee Howard, the chief executive officer of Brandthropologie, suggested that the word *collaboration* is more appropriate (Rowe, 2016). Nevertheless, Meelen and Frenken (2015) argued that if the hosts rent out their home while staying temporarily elsewhere due to, for example, vacation, business trip and family visit, then Airbnb in this case is part of sharing economy as the hosts' otherwise underutilized house is now being shared. However, if the hosts live permanently in another house and continuously rent out their own house, Airbnb in this case is not part of the sharing economy. In line with this view, we consider Airbnb part of the sharing economy under an assumption that Airbnb's hosts are not running a (often illegal) hotel by permanently living in another house.

Researchers described Airbnb as a disruptive business model, posing threats to the traditional hospitality industry (Guttentag, 2015; Geissinger et al., 2018). The initial market of Airbnb was limited in size, so it was unappealing to attract attention from leading companies. However, over time Airbnb improves and attracts an increasing number of customers, making leading companies start to view competition from Airbnb and other similar platforms as a serious threat. On the other hand, facing regulations from cities around the world, along with competition from other companies in the industry, Airbnb has to rethink its business (Kerr, 2018). To maintain its competitiveness, Airbnb has continually introduced noteworthy service improvement. For example, in March 2018, it released 21 new accessibility filters, such as step-free entry to rooms and entryways that are wide enough to accommodate a wheelchair, across the platform that make it easier for guests with disabilities to find accessible travel accommodation worldwide (Airbnb, 2018c). It will also expand Experiences (excursions or other activities designed and led by local hosts) to 1000 destinations by the end of 2018, including places such as Easter Island, Tasmania and Iceland (Airbnb, 2018d). Its revenue represented a significant fraction of hotels (Zervas et al., 2017). Unlike most platforms in the sharing domain, Airbnb has achieved profitability in 2016 (Lutz & Newlands, 2018).

Disruptive ideas like Airbnb have the potential to change the way an industry operates, and the success of Airbnb confirmed that once the change is initiated, it is highly unlikely that the industry would revert to the old model (Varma et al., 2016). Therefore, studies regarding Airbnb deserve researchers' attention. While there are prior studies studying Airbnb's impacts to the hotel industry (e.g. Oskam & Boswijk, 2016; Xie & Kwok, 2017; Zervas et al., 2017; Blal et al., 2018) and its role as disruptive innovation (e.g. Guttentag, 2015; Guttentag & Smith, 2017; Geissinger et al., 2018), there is a surprising dearth of research regarding the discovery of knowledge from the sharing platform in order to understand consumer behaviour.

2.2 Consumers' expectation on Airbnb

The Internet has created new ways of sharing as well as facilitating older forms of sharing on a larger scale (Belk, 2014b). In general, trust plays an important role in decision making in online transactions (Ruan & Durresi, 2016). Therefore, many online platforms provide consumer reviews so as to allow consumers gather more information for decision making. Consumer reviews are often provided in two formats: average ratings giving an overview over the overall perceived quality of the product (i.e., statistical information) and single reviews that contain personal narratives of experiences made with a specific product (Helversen et al., 2018). In addition, reviews help consumers assess risks before agreeing to a transaction. Compared with traditional hotels that reduce risks through standardisation, safety regulations and business reputation, Airbnb has a higher level of risks as it involves admitting strangers to one's private environment. The online review system is considered the foundation of mutual trust between hosts and guests in Airbnb (Chen & Chang, 2018). All reviews on Airbnb are written by authenticated hosts and guests after the stay and both parties are not allowed to remove reviews unless the reviews violate Airbnb's content policy.

Previous research studies were conducted to examine Airbnb's consumers' expectation. For instance, Guttentag (2015) suggested that three important Airbnb's distinct appeals are price, amenities and authenticity. Wang and Nicolau (2017) examined Airbnb price determinants from five categories that are host, site and property, facility and service, rental rules and online review score. Ju et al. (2018) discovered that major factors affecting Airbnb's customer satisfaction are host, room/house, location and neighbourhood. Cheng and Jin (2019) identified that key attributes influencing Airbnb users' experiences include location, amenities and host.

While relevant existing work serves as useful references to understand consumers' expectation on Airbnb's services, the findings vary with the extant studies. Many of the studies, as shown in Table 1, were conducted through text mining from a limited size of samples, or by analysing primary data collected from surveys and/or interviews. The motivations of this study are twofold. First, we expect that Airbnb will continue its rapid growth and attract more users worldwide. The profitability achieved by Airbnb also proved the inception of peer-to-peer accommodation in the global market. Therefore, studies regarding Airbnb deserve research attention. Second, to our knowledge, what is missing thus far in the growing literature on the sharing economy, and on Airbnb in particular, is the examination of change in consumers' expectation across the timespan (i.e. longitudinal analysis) and seasons (i.e. seasonal analysis). This study aims to analyse a big data set of Airbnb's online review comments in an attempt to understanding and detecting any changes in consumers' expectation by year and by season.

Table 1. Description of the data set used in 7 similar papers

Reference	Sample size	Data type	Source(s)	Geographic scope	Longitudinal analysis?	Seasonal analysis?
Varma et al. (2016)	347 12	Primary data	Survey Interview	Not defined	×	×
Bae et al. (2017)	411	Primary data	Survey	South Korea	×	×
Brochado et al. (2017)	1,776	Secondary data	Online review	India, Portugal and the U.S.	×	×
Wang & Nicolau (2017)	298,331	Secondary data	Airbnb's listing	33 cities	×	×
Blah et al. (2018)	11	Secondary data	Airbnb's listing	San Francisco	×	×
Ju et al. (2018)	16,340 322	Secondary data Primary data	Online review Survey	Miami, New York, San Francisco and Chicago	×	×
Cheng & Jin (2019)	181,263	Secondary data	Online review	Sydney	×	×
This study	169,666	Secondary data	Online review	London	✓	✓

3. Methodology

To ensure the generalisation of the framework, the design of the methodology is derived from previous studies. Based on the methodology developed by Vecchio et al. (2018) and Cheng and Jin (2019), the two key phases for analysing data, such as online reviews from social media, are data collection and data analysis. In the data analysis phase, textual data has to pre-processed to obtain a noiseless and clean dataset. In line with this, we create an additional phase, namely Data Pre-processing phase, to emphasize the text pre-processing tasks as the data involved in this study are all textual comments. Furthermore, visualisation is found helpful in getting first-hand insights based on the analysis results (Mahgoub, et al., 2008; Kulkarni & Kulkarni, 2016). Thus, appropriate visualisation tools are included in the proposed data analysis phase. Consequently, the methodology, as shown in Figure 1, can be divided into three phases which are Data Collection, Data Pre-processing, and Data Analysis. In Data Collection, raw unstructured data from Airbnb's website is collected. Data includes attributes related to Airbnb's listings, guests' textual reviews and hosts' profiles. After the data is collected and aggregated, the dataset is further split into sub-datasets according to years and seasons. In Data Pre-processing, the data is converted to a format that is appropriate for data analysis. In particular, the unstructured text needs to be converted into a semi-structured dataset so that one can apply analytics techniques to discover hidden knowledge in the next phase, i.e. Data Analysis, where the hierarchical clustering method is used to assign keywords into a group of clusters. Details of each phase are described in the following sections.

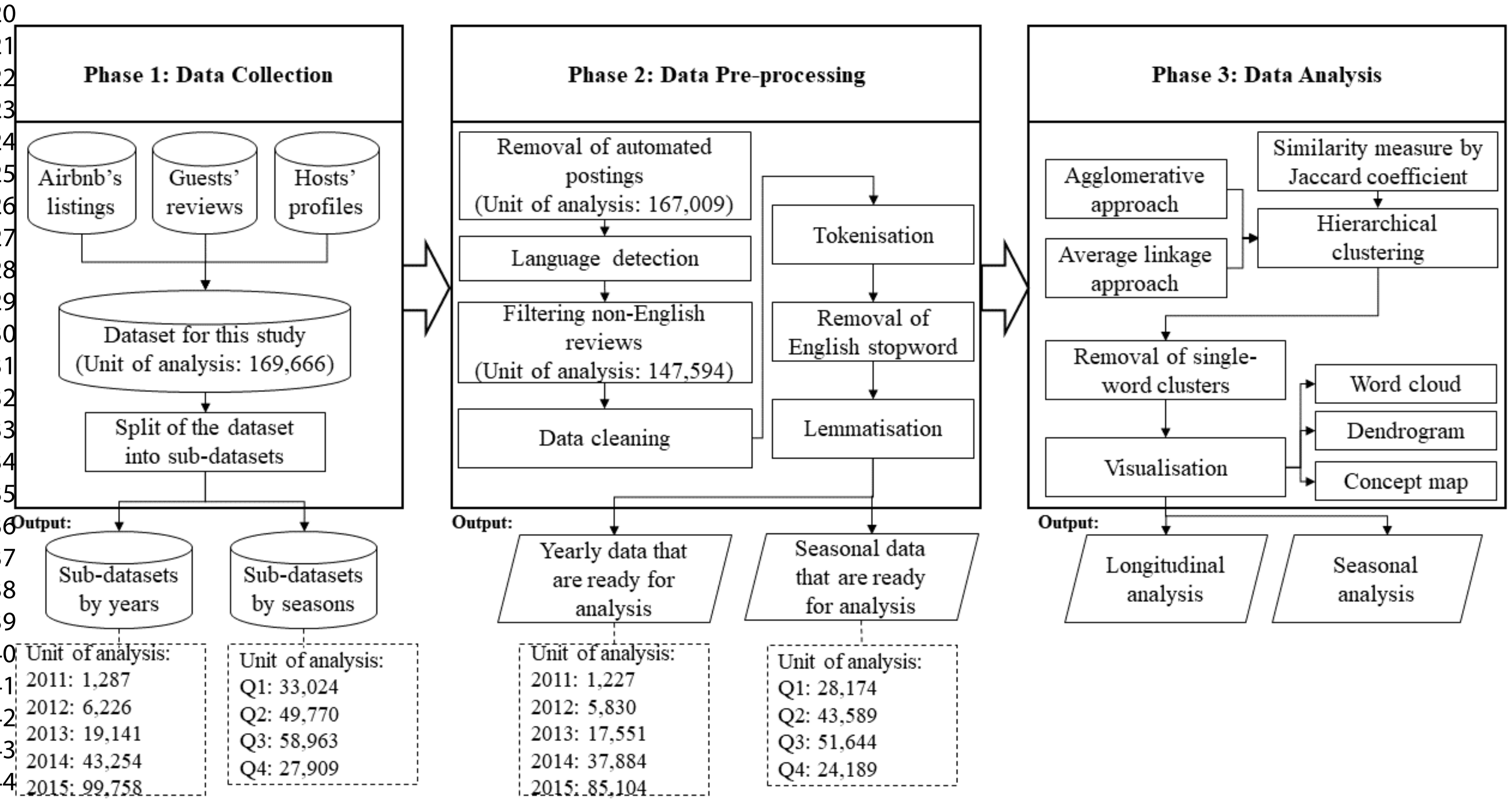


Figure 1. Methodology of this study

3.1 Data collection

The original dataset contains 169,666 records related to Airbnb's listings located in London from 2011 to 2015. Each record consists of attributes including review id, reviewer id, reviewer name, host id, host name, review date and reviews posted by reviewers, i.e. the guests who stayed in the listings provided by the hosts. The reviews are textual comments written in different languages such as English, Spanish, Chinese and Korean. Each of them is a 500-character-limit description of the guest's experience that could be positive, negative or neutral. The output of this phase is sub-datasets split according to years and seasons. For longitudinal analysis, the dataset is divided into 5 sub-datasets by years from 2011 to 2015. For seasonal analysis, the dataset is divided into sub-datasets in four quarters per year based on the review date. Quarter 1 (Q1) is from 1 March to 31 May, i.e. spring in London. Quarter 2 (Q2) is from 1 June to 31 August, i.e. summer in London. Quarter 3 (Q3) is from 1 September to 30 November, i.e. autumn in London. Quarter 4 (Q4) is from 1 December to 29 February, i.e. winter in London. The size of each sub-dataset is shown in Table 2. The total number of reviews increases with year from 2011 to 2015. This indicates that Airbnb is gaining its popularity among tourists who visit London. Furthermore, the number of comments posted in Q2 and Q3 are larger than that in Q1 and Q4. This shows that there are more tourists visiting London during the period from June to November.

Table 2. Size of data collected for this study

Year	Q1	Q2	Q3	Q4	Total
2011	211	356	514	206	1,287
2012	1,347	1,549	2,405	925	6,226
2013	3,779	5,818	6,494	3,050	19,141
2014	8,425	12,770	14,888	7,171	43,254
2015	19,262	29,277	34,662	16,557	99,758
Total	33,024	49,770	58,963	27,909	169,666

3.2 Data pre-processing

The dataset contains records where the guests or hosts cancel the reservation before the stay. In those records, the reviews are automatically generated by Airbnb. There are three types of automated postings found: "*The reservation was canceled X day(s) before arrival. This is an automated posting*", "*The host canceled this reservation X day(s) before arrival. This is an automated posting*" and "*Tell others in the Airbnb community about your stay*". As these automated postings provide no insights on guest's experience, they are removed from our dataset. After omitting these automated postings, the dataset size is reduced to 167,009.

The 500-character limit of reviews makes it feasible for Airbnb users to post comments in multiple languages. This practice is common in the Airbnb review system because users are worldwide, speaking in different languages. Therefore, language detection and filtering is a critical step before modelling. After filtering non-English reviews, the dataset size is further reduced to 147,596 as shown in Table 3. Furthermore, when the seasonal analysis is conducted, the model takes the review dates as the input attribute. Data cleaning is needed to correct any inconsistent dates appeared in the reviews. When users rate their stay experience, it is observed that they might express the ratings in the form of X/Y (e.g. 5/5, implying 5 out of 5). Nevertheless, this kind of expression can be mistaken as dates (e.g. 5/5 and 10/10 could be mistaken as 5th May and 10th October, respectively) in the reviews and create inconsistencies during the seasonal analysis when the correct data format is missing.

After data cleaning, tokenisation is performed to convert each word in a review into a distinct attribute (Kayser & Blind, 2017). Figure 2 shows the number of tokens created in the yearly-segmented reviews. After that, Standard English stopword filtering is applied, followed by lemmatisation that works on a bunch of rules where the basic idea is to covert different inflected forms of a work into a single item to be analysed.

Table 3. Size of data written in English only

Year	Q1	Q2	Q3	Q4	Total
2011	207	340	480	200	1,227
2012	1,273	1,454	2,244	859	5,830
2013	3,395	5,397	5,962	2,797	17,551
2014	7,391	11,246	12,950	6,297	37,884
2015	15,908	25,152	30,008	14,036	85,104
Total	28,174	43,589	51,644	24,189	147,596

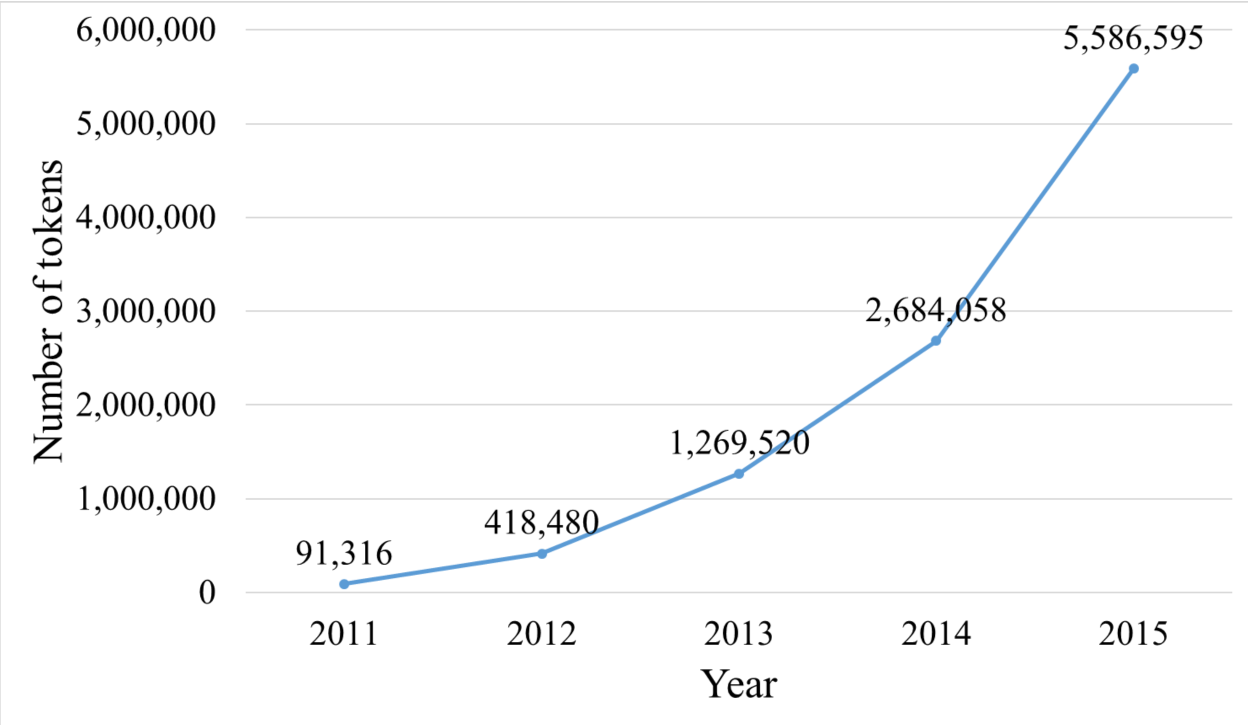


Figure 2. Number of tokens in yearly-segmented reviews

3.3 Data analysis

This phase generates results of the longitudinal analysis and the seasonal analysis. In the former analysis, text mining is applied to each year's dataset to see how themes in the reviews change over the 5-year timespan. In the latter analysis, the dataset is split into multiple sub-datasets based on seasons. The motivation is to investigate if guests' expectation changes when they visit London in different climates. To achieve this, the text mining software QDA Miner is used to find groups of similar reviews in a collection of reviews. It uses a hierarchical clustering method to build a group of clusters while the hierarchy is constructed using the agglomerative approach. In the agglomerative approach, words that tend to appear together are combined at an early stage while those that are independent from one another or those that don't appear together tend to be combined at the end of the agglomeration process (Allahyari et al., 2017). There are three different merging methods in agglomerative approaches, single linkage method, average linkage method and complete linkage method. In the single linkage method, the similarity between two clusters is the highest similarity between any pair of items in these groups. In the average linkage method, the similarity between two clusters is the average similarity between pairs of items in these groups. In the complete linkage method, the similarity between two clusters is the worst-case similarity between any pair of items in these groups. In QDA Miner, the average-linkage method is used.

There are various measures to compare the similarities of items. One of the selection criteria is the binary symmetry. Some measures such as the simple matching coefficient consider 0-0 matches given that there is no clear asymmetry between group 0 and group 1. On the other hand, some measures such as Jaccard coefficient ignore 0-0 matches given that 0-0 matches are uninformative. In this study, Jaccard coefficient is selected because not counting 0-0 matches is important or most of the words would be found highly similar to most of the other words. It measures the similarity between words *i* and *j*, as in Eq. (1):

$$sim_{Jaccard}(i,j) = \frac{a}{a+b+c} \tag{1}$$

where *a* is the number of cases where both *i* and *j* are present, *b* is the number of cases where only *i* is present and *c* is the number of cases where only *j* is present. It assigns equal weight to matches and non-matches, but zero weight to 0-0 matches. If the Jaccard coefficient is equal to 0, it means that the two words have no similarity. On the other hand, if the Jaccard coefficient is equal to 1, it means that the two words are identical matches.

After the computation of similarity, visualisation is applied to let the data analysts see underlying patterns in the datasets. Techniques used in this study include word clouds, dendrograms and concept maps. Examples of these visualisation techniques are given in Section 4. Compared with mining other social data such as Twitter data that are only short posts or tweets within 280-character limit, mining Airbnb data involves a higher level of complexity as each review can be up to 500 characters. Besides, the low standardisation of Airbnb services delivered by non-professional individual hosts makes every stay experience very unique. As such, topics covered in the Airbnb reviews are in a high variety, making keyword clustering more computationally expensive. For instance, a very long dendrogram including many clusters of isolated items could be resulted. In view of this, the design of the analytical model involves single-word cluster removal. We simplify the use of the dendrogram by hiding all single item clusters and allowing one to concentrate only on the strongest associations.

4. Results and Discussion

To perform text mining, the databset is divided into sub-datasets according to year and season for longitudinal analysis and seasonal analysis, respectively. After that, each dataset is inputted into QDA Miner. After data pre-processing as stated in Section 3, a list of words is generated. The frequency and Term Frequency–Inverse Document Frequency (TF-IDF) of each word is computed as shown in Table 4. Word clouds are used as a graphic representation of the frequency of the words. An example of a word cloud is shown in Figure 3. The size of the words indicates the frequency of the words. With the use of word clouds, managers can have a quick view to understand the keywords that are frequently mentioned by customers. It is an efficient tool for prioritising customer concerns to design and improve services offerings.

The average-linkage hierarchical clustering results are presented in the form of a dendrogram as shown in Figure 4. The vertical axis is made up of the items and the horizontal axis represents the clusters formed at each step of the clustering procedure. In general, there are two types of keyword clustering, first order clustering and second order clustering. The former one is based on keyword co-occurrences and will group together words appearing near each other or in the same document. The latter one is based on co-occurrence profiles and will consider that two keywords are close to each other, not necessarily because they co-occur but because they both occur in similar environment. Second order clustering is chosen for this study as it has the capability in grouping words that are synonyms or alternate forms of the same word. For example, while *Underground* and *Tube* will seldom or never occur together in the same document, second order clustering may find them very close because they both co-occur with words such as *Station*. It is also able to group words that are related semantically such as *Apartment*, *Flat*, *Room*, and *House*, because of their propensity to be associated with similar verbs like *Stay*.

Jaccard coefficient is used to compute similarity and concept maps, as depicted in Figure 5, are graphic representation of the proximity values computed on the keywords using multi-dimensional scaling. In a concept map, a node represents a keyword and its size indicates the frequency of the keyword (Tse et al., 2016). The distances between pairs of keywords indicate how likely those keywords are to appear together. Keywords that appear close together in the map usually tend to occur together while keywords that are independent from one another or that seldom appear together are located far from each other. The colour of the nodes indicates the keyword's membership to different clusters, each of which represents a discussion theme. The concept maps are useful in detecting hidden knowledge that may explain similarities between keywords. Customers may express similar comments by using similar words but not the exact sentence. In this case, managers can study the similarities of the words by referring to the node distributions. This approach is particularly useful to explore the customers' perception of some important topics or themes in which the managers may overlook some wordings which are co-occurrences in the comments. Nevertheless, a limitation of the concept maps is that clear visualisation is enabled only when the number of nodes is limited. When there is a large number of nodes shown in the concept map, it becomes difficult for managers to have a quick overview of the co-words. Another limitation is that it is not possible to position nodes in a two-dimensional concept map in a way that the distance between any pairs of nodes can reflect their similarities with 100% accuracy. Therefore, it is worth noting that only the approximate similarities between words are shown in a two-dimensional concept map.

Existing literature identified that tourists generally use similar sets of attributes to evaluate their accommodation experiences. For example, common attributes include location, amenities and price. However, for Airbnb services, there are additional important attributes such as host, cleanliness and homeliness. Common and additional attributes across the 5-year timespan are highlighted in Figure 6. In our longitudinal analysis, it is found that Airbnb's guests value the helpfulness and flexibility of the host as well as the communication with the hosts. The experiences evaluated by the guests are not only limited to how they feel *during* the stay, but can also be related to the effectiveness of communication with the hosts *before* the stay as well as the helpfulness of the hosts *after* the stay when help is needed (e.g. when guests left their belongings in the place after checking out). Furthermore, it is worth noting that a number of guests visiting London stated that they visited Museums. They appreciated if hosts provided them with advice on sightseeing, in particular, related to museums and theatres.

Table 4. Frequency and TF-IDF of the top ten keywords

	Frequency	% Shown	% Processed	% Total	No. of Cases	% Cases	TF - IDF
LONDON	39175	2.05%	1.45%	0.70%	29625	34.81%	17953.6
FLAT	34394	1.80%	1.27%	0.62%	21550	25.32%	20516.1
PLACE	33473	1.75%	1.24%	0.60%	25026	29.41%	17792.9
ROOM	32827	1.72%	1.22%	0.59%	25830	30.35%	16998.7
HOST	32061	1.68%	1.19%	0.57%	29569	34.74%	14719.6
CLEAN	31205	1.63%	1.16%	0.56%	29342	34.48%	14431.1
APARTMENT	30110	1.58%	1.12%	0.54%	19344	22.73%	19372.9
NICE	29606	1.55%	1.10%	0.53%	23650	27.79%	16464.5
LOCATION	28814	1.51%	1.07%	0.52%	27231	32.00%	14259.7
RECOMMEND	24672	1.29%	0.91%	0.44%	24198	28.43%	13475.1

Figure 3. An example of a word cloud

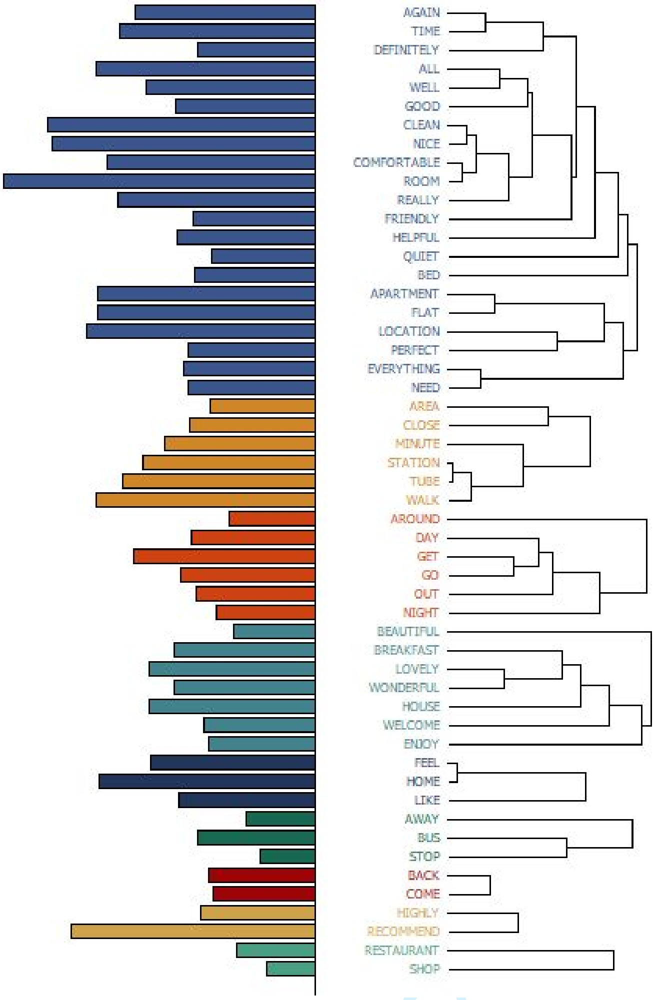


Figure 4. An example of a dendrogram

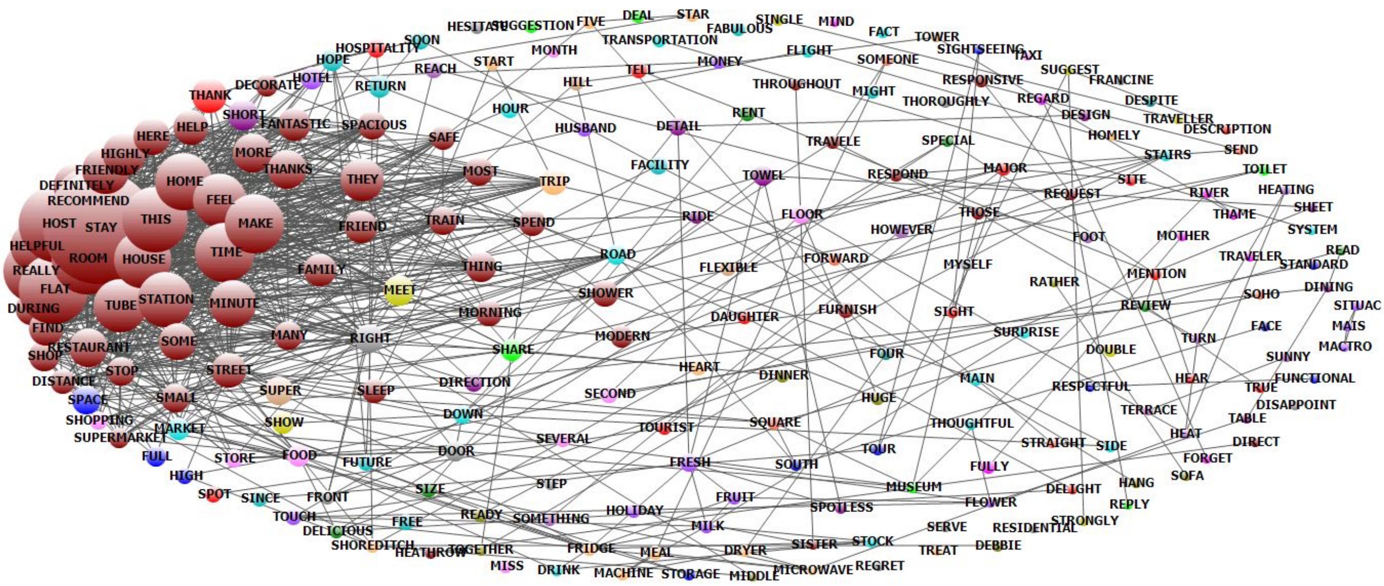


Figure 5. An example of a concept map of keywords using hierarchical clustering

Compared with traditional hotels that reduce risks through standardisation, safety regulations and business reputation, Airbnb has a higher level of risks as it involves admitting strangers to one's private environment. We expected that safety would be one of the attributes highlighted in the reviews (Huber, 2017; Karlsson et al., 2017; Alrawadieh & Alrawadieh, 2018). However, surprisingly, safety is less frequently mentioned in the reviews compared with cleanliness and homeliness. Though there exist some reviews complaining about the unsafe neighbourhood, guests generally treat this minor as long as the apartment they stay is clean or feels like home to them. On the other hand, events, such as major sports or cultural events, generate tourism activities and accommodation located close to the events is more expensive due to the high demand (Herrmann & Herrmann, 2014; Tussyadiah, 2016; Fiarley & Dolnicar, 2017). We thus expected that there would be more reviews highlighting the convenience of locations for connections to Royal Wedding location in 2011. Nevertheless, there are not many reviews related to stay experience during the Royal Wedding week.

Furthermore, noise is found to be one of the most frequently appeared words in the reviews posted in 2012 and 2013. This finding is quite interesting since this is not a common wording appeared in other hotel reviews. A number of guests stated that they had negative feelings during the stay due to noises caused by construction work, traffic, and bars nearby. Another interesting finding is, from 2011 to 2015, there is an increasing number of guests viewing internet access as a major issue. Positive review comments associated with internet access include stable and fast Wi-Fi connection. Some guests compare the Wi-Fi connection in Airbnb's listings with that in hotels. This indicates the failure of Wi-Fi connection in Airbnb that can ruin the entire staying experience.

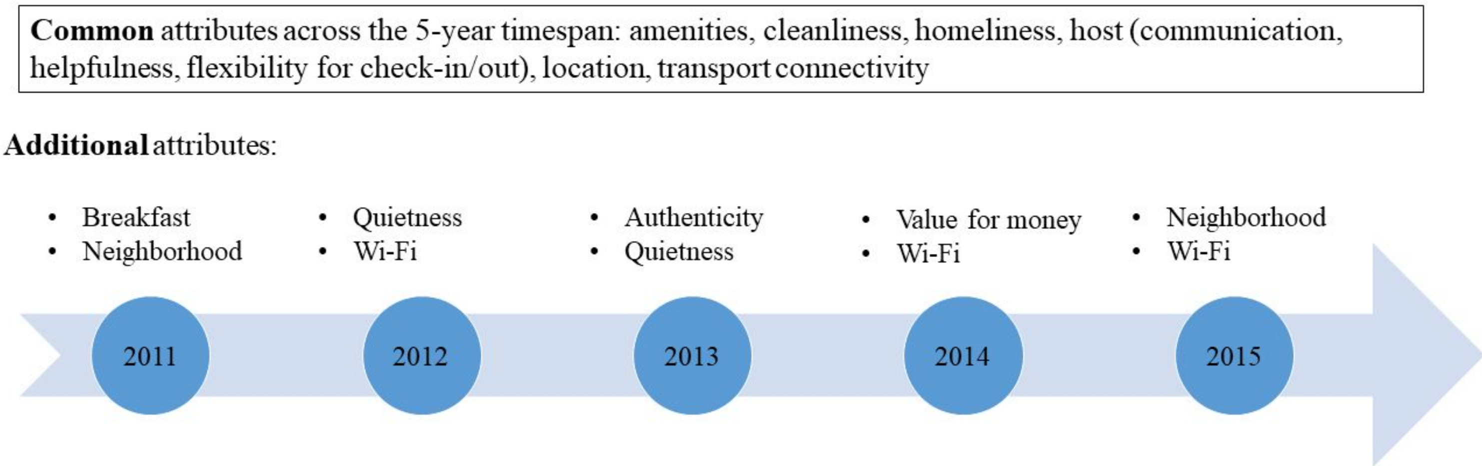


Figure 6. Common and additional attributes across the 5-year timespan

From the seasonal analysis, the results indicate that guests value the neighbourhood more in summer than in other seasons. A number of guests view the location as a better choice if it is near parks with ponds for swimming and space for sunbathing in summer. Meanwhile, guests highlight the importance of having a washing machine in the place. This is particular true to the guests who have a relative long stay during spring and summer in London. Guests prefer fresh clothes because of the relatively long daylight hours in spring and summer.

On the contrary, during winter, guests believe that being able to enjoy hot shower in a clean bathroom is a big advantage. Negative comments associated with amenities in winter include not enough hot water for shower, doors/windows had to be closed to keep warm, and limited hours for heating per day. Interestingly, fast internet connection and television are important attributes to guests in winter. This could be because guests tend to spend more time indoor because of the low temperature in winter and thus amenities providing entertainment become crucial to them.

Based on the abovementioned observations, recommendations that can help the hosts improve their services are summarised in Table 5.

Table 5. Recommendations for hosts in different seasons

Season	Recommendations for hosts
Q1	Highlight if the listing has an in-house washing machine (that allows guests to use) or any laundry services nearby so that guests can keep their clothes clean; provide clear instructions on how to use the washing machine.
Q2	The listing is more attractive to guests if guests can enjoy sunbathing or swimming nearby.
Q3	*[Q3 does not have particular discussion themes]*
Q4	Highlight the amenities such as hot water that can give guests a warm and cosy stay; provide fast internet connection and TV so that guests can entertain when indoor.

To summarise, the results suggest that although Airbnb customers might use similar attributes associated with hotel stays to evaluate their accommodation experience, the priorities of these attributes might vary in years and seasons. The change in customer expectation is due to the low predictability of the Airbnb services. There are two possible explanations. First, compared to hotel services, Airbnb services are not standardised as the service delivery process is done by individual hosts, instead of professional providers (Tussyadiah & Zach, 2017), leading to both good and bad experiences. Second, the evaluation of accommodation services is subjective in nature, depending on individual guests. For instance, while some researchers stated that having a chance to build friendship with local hosts makes customers prefer Airbnb to hotel services, it happens that some customers reported negative experiences when the hosts spent too much time in the place during their stay. Despite the fact that Airbnb experience is less predictable, keywords that discovered from this study serve as useful references for both Airbnb and individual hosts to design and improve services. The results provide additional insights for more proactively aligning services with customer expectations when the time element is considered.

5. Conclusion and Future Work

The rapid development of information technologies has allowed people to share their goods with strangers online. As a typical example of the sharing economy, Airbnb connects hosts and guests by sharing part or all of homes as rental properties for short stay. In general, people are wary of hosting a stranger or sleeping in a stranger's home. To deal with the online trust problem, Airbnb users are encouraged to share their Airbnb experiences by posting public reviews online after each stay. The purpose is to allow people to view other users' comments before agreeing to any transaction. Prior studies confirmed that consumer expectation of the stay comes from the earlier shared experiences they found in the reviews. Analysing online reviews can thus uncover important service attributes that influence accommodation booking intention.

This paper presents a methodology for mining online reviews to discover Airbnb customer needs. It provides implications for the sharing economy to prioritise customer concerns to design and improve service offerings. From the perspective of hosts, the knowledge discovered from this empirical study provides decision support for them to more proactively improve and align services with the guests' expectation. Being able to come up with new or improved services is important to generate future customer requests (Lee, 2018). From the perspective of Airbnb, the empirical results provide Airbnb with additional insights on criteria that guests consider the most, and those criteria are useful in improving the features of the websites. A better match between the hosts and the guests can be achieved and thus increases the profit of Airbnb as Airbnb charges a service fee when a booking is confirmed. Lastly, from the perspective of guests, their experience in using the Airbnb services can be improved when the Airbnb's online platform provides better filters for their search and the hosts can sufficiently meet their expectations during their stay. This can facilitate the engagement with the sharing economy in the long term. In general, this methodology can be applied in other peer-to-peer service platforms, such as eBay, where two individuals can interact directly with each other without intermediation by a third-party. For this kind of platforms, online review systems appear to be the foundation of mutual trust between sellers and buyers. The proposed methodology is useful in analysing the online reviews posted by users and discovering insights for service offering improvements. Though the dataset used in this study is from 2011 to 2015 that might be not very recent, this study provides a foundation for analysing social big data in the sharing economy that has a huge potential to grow. In addition, it is of high potential generalisation for applications in tourism-related industries such as the niche tourism where the key to success depends on how service providers or enablers satisfy specific needs of customers (Wu et al., 2016). The proposed framework can be applied to generate distinctive tourism solutions that align with customers' interests. By identifying the key factors affecting tourism performance and creating more personalised offerings, tourism industrialists can strengthen their competitive edge (Do & Chen, 2013; Vecchio et al., 2018; Wu et al., 2018).

This paper also provides practical implications for mining social data that are in high variety in terms of data length and topics. Key challenges lie in data pre-processing and data analysis. Of special note during data pre-processing is that multiple languages may occur per review. Thus, analysts have to detect languages and decide the filtering criteria when multiple languages per review occurs. Moreover, analysts have to be aware that the similarity defined by clustering depends strongly on how the distance measure captures the concept of dissimilarity. The interpretation of dendrograms as well as concept maps can be, in several domains, quite subjective.

This paper conducts both longitudinal analysis and seasonal analysis of Airbnb customer behaviour. Our findings reveal how users' mind-set of good quality of accommodations changes over a 5-year timespan and in different quarters throughout the years. This is considered a significant contribution as, to our knowledge, what is missing thus far in the existing literature on the sharing economy, and on Airbnb in particular, is the examination of change in consumers' expectation across the timespan and seasons. Some may argue that our findings that are obtained based on the historical data from 2011 to 2015 may no longer be valid in the future because the global market in the hospitality industry is very dynamic. However, we believe that some of the findings discovered in this study will remain important for shaping the sharing-accommodation services in the future. The key service attributes identified in this study may eventually become basic requirements of the customers. For instance, from the longitudinal analysis, it is found that there is an increasing trend of requesting fast internet connection in Airbnb's listings. As the technology is getting more advanced, it is believed that wireless internet connection will become a fundamental attribute in the future. This study is able to highlight some upcoming trends that should not be overlooked in the future. In addition, the results from the seasonal analysis will still be valid as we expect that the seasonal weather in London should be similar every year. Thus, the recommendations made in this study are worth to be considered if one aims to improve the accommodation service of a listing in different seasons.

This study also opens up a number of future research avenues. First, testing the causal relationship among meta-data and quantifying the attributes in textual data will be a useful next step forward, providing additional insights for the sharing economy. To achieve this, reviews are suggested to be linked with the corresponding hosts and listings, followed by a regression analysis to study whether hosts' attributes (e.g. age, gender) and listings' attributes (e.g. price, location) are associated with the consumer behaviour as explained by each theme. Second, considered that hosts are encouraged to post their personal photos together with their listings, it is suggested that personal photos of Airbnb's hosts can be analysed. Based on the photos, age, gender as well as emotions of hosts can be extracted as attributes using face recognition technologies. Last but not least, while this paper is confined to Airbnb's listings located in London, it is recommended that researchers apply the methodology to datasets of Airbnb's listings in different cities. It would be interesting to investigate any change in behaviour of geographically dispersed customers.

References:

Airbnb (2018a). *About Us.* https://press.atairbnb.com/about-us/ (Accessed 10 August 2018).

Airbnb (2018b). *What is the Airbnb service fee?* https://www.airbnb.co.uk/help/article/1857/what-is-the-airbnb-service-fee (Accessed 10 August 2018).

Airbnb (2018c). *Airbnb Highlights New Accessibility Filters and Features for Guests with Disabilities Worldwide.* https://press.atairbnb.com/airbnb-highlights-new-accessibility-filters-and-features-for-guests-with-disabilities-worldwide/ (Accessed 13 August 2018).

Airbnb (2018d). *Airbnb Doubles Down on Experiences, Expanding to 1000 Destinations and Adding New Passion Categories in 2018.* https://press.atairbnb.com/airbnb-doubles-down-on-experiences-expanding-to-1000-destinations-and-adding-new-categories-in-2018/ (Accessed 13 August 2018).

Allahyari, M., Pouriyeh, S.A., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B., & Kochut, K.J. (2017). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *CoRR, abs/1707.02919*.

Alrawadieh, Z., Alrawadieh, Z. (2018). Exploring entrepreneurship in the sharing accommodation sector: Empirical evidence from a developing country. *Tourism Management Perspectives 28*, 179-188.

Bae, S.J., Lee, H., Suh, E.-K., Suh, K.-S. (2017). Shared experience in pretrip and experience sharing in posttrip: A survey of Airbnb users. *Information & Management 54*(6), 714-727.

Bardhi, F., Eckhardt, G.M. (2012). Access-Based Consumption: The Case of Car Sharing. *Journal of Consumer Research 39*(4), 881-898.

Basili, R., Croce, D., Castellucci, G. (2017). Dynamic polarity lexicon acquisition for advanced Social Media analytics. *International Journal of Engineering Business Management 9*, 1-18.

Belk, R. (2014a). Sharing Versus Pseudo-Sharing in Web 2.0. *Anthropologist 18*(1), 7-23.

Belk, R. (2014b). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research 67*(8), 1595-1600.

Belk, R.W. (2013). Extended Self in a Digital World. *Journal of Consumer Research 40*(3), 477-500.

Blal, I., Singal, M., Templin, J. (2018). Airbnb's effect on hotel sales growth. *International Journal of Hospitality Management 73*, 85-92.

Botsman, R. (2013). *The Sharing Economy Lacks A Shared Definition.* https://www.fastcompany.com/3022028/the-sharing-economy-lacks-a-shared-definition, Retrieved 27 July 2018.

Brochado, A., Troilo, M., Shah, A. (2017). Airbnb customer experience: Evidence of convergence across countries. *Annals of Tourism Research 63*, 210-212.

Chen, C.-C., Chang, Y.-C. (2018). What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness. *Telematics and Informatics 35*(5), 1512-1523.

Cheng, M., Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management 76*(Part A), 58-70.

Cohen, B., Kietzmann, J. (2014). Ride On! Mobility Business Models for the Sharing Economy. *Organization & Environment 27*(3), 279-296.

Do, Q.H., Chen, J.F. (2013). Prioritizing the factor weights affecting tourism performance by FAHP. *International Journal of Engineering Business Management 5*, 1-10.

Fairley, S. and Dolnicar, S. (2017) Chapter 10 – Supporting Events, in S. Dolnicar, *Peer-to-Peer Accommodation Networks: Pushing the boundaries*, Oxford: Goodfellow Publishers, pp. 109–119, https://dx.doi.org/10.23912/9781911396512–3608

Gavilan, D., Avello, M., Martinez-Navarro, G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management 66*, 53-61.

Geissinger, A., Laurell, C., Sandström, C. (2018). Digital Disruption beyond Uber and Airbnb – Tracking the long trail of the sharing economy. *Technological Forecasting & Social Change*. https://doi.org/10.1016/j.techfore.2018.06.012.

Gunter, U. (2018). What makes an Airbnb host a superhost? Empirical evidence from San Francisco and the Bay Area. *Tourism Management 66*, 26-37.

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector, *Current Issues in Tourism 18*(12), 1192-1217.

Guttentag, D.A., Smith, S.L.J. (2017). Assessing Airbnb as a disruptive innovation relative to hotels: Substitution and comparative performance expectations. *International Journal of Hospitality Management 64*, 1-10.

Hamari, J., Sjöklint, M., Ukkonen, A. (2016). The Sharing Economy: Why People Participate in Collaborative Consumption. *Journal of the Association for Information Science and Technology 67*(9), 2047-2059.

Heinrichs, H. (2013). Sharing Economy: A Potential New Pathway to Sustainability. *GAIA 22*(4), 228-231.

Herrmann, R., Herrmann, O. (2014). Hotel roomrates under the influence of a large event: The Oktoberfest in Munich 2012. *International Journal of Hospitality Management 39*, 21-28.

Huber, A. (2017). Theorising the dynamics of collaborative consumption practices: A comparison of peer-to-peer accommodation and cohousing. *Environmental Innovation and Societal Transitions 23*, 53-69.

Jøsang, A., Ismail, R., Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision Support Systems 43*(2), 618-644.

Ju, Y., Back, K.-J., Choi, Y., Lee, J.-S. (2018). Exploring Airbnb service quality attributes and their asymmetric effects on customer satisfaction. *International Journal of Hospitality Management*, https://doi.org/10.1016/j.ijhm.2018.07.014

Karlsson, L., Kemperman, A., Dolnicar, S. (2017). May I sleep in your bed? Getting permission to book. *Annals of Tourism Research 62*, 1-12.

Kayser, V., Blind, K. (2017). Extending the knowledge base of foresight: The contribution of text mining. *Technological Forecasting and Social Change 116*, 208-215.

Kerr, D. (2018). *Aribnb will expand 'experiences' to 1,000 cities this year*. https://www.cnet.com/news/airbnb-rolls-out-experiences-to-1000-new-cities/ (Accessed 13 August 2018).

Kulkarni, M., Kulkarni, S. (2016). Knowledge Discovery in Text Mining using Association Rule Extraction. *International Journal of Computer Applications 143*(12), 30-35.

Lee, C.K.H. (2018). A review of applications of genetic algorithms in operations management. *Engineering Applications of Artificial Intelligence 76*, 1-12.

Lee, E.-J., Shin, S.Y. (2014). When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo. *Computers in Human Behavior 31*, 356-366.

Lutz, C., Newlands, G. (2018). Consumer segmentation within the sharing economy: The case of Airbnb. *Journal of Business Research 88*, 187-196.

Mahgoub, H., Rösner, D., Ismail, N., Torkey, F. (2008). A Text Mining Technique Using Assocition Rules Extraction. *International Journal of Computer and Information Engineering 2*(6), 2044-2051.

Meelen, T., Frenken, K. (2015). Stop Saying Uber Is Part of The Sharing Economy: What is beng shared being your money? https://www.fastcompany.com/3040863/stop-saying-uber-is-part-of-the-sharing-economy (Accessed 12 August 2018).

Oskam, J., Boswijk, A. (2016). Airbnb: the future of networked hospitality businesses. *Journal of Tourism Futures 2*(1), 22-42.

Puschmann, T., Alt, R. (2016). Sharing Economy. *Business and Information Systems Engineering 58*(1), 93-99.

Rowe, D.S. (2016). The Rise of the Sharing Economy. *Customer Relationship Management 20*(10), 22-25.

Ruan, Y., Durresi, A. (2016). A survey of trust management systems for online social communities – Trust modeling, trust inference and attacks. *Knowledge-Based Systems 106*, 150-163.

Sparks, B.A., Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management 32*(6), 1310-1323.

Troisi, O., D'Arco, M., Loia, F., Maione, G. (2018). Big data management: The case of Mulino Bianco's engagement platform for value co-creation. *International Journal of Engineering Business Management 10*, 1-8.

Tse, Y. K., Zhang, M., Doherty, B., Chappell, P., Garnett, P. (2016). Insight from the horsemeat scandal: exploring the consumers' opinion of tweets toward Tesco. *Industrial Management & Data Systems 116*(6), 1178-1200.

Tussyadiah, I.P. (2016). Factors of satisfaction and intention to use peer-to-peer accommodation. *International Journal of Hospitality Management 55*, 70-80.

Tussyadiah, I.P., Pesonen, J. (2016). Impacts of peer-to-peer accommodation use on travel patterns. *Journal of Travel Research 55*, 1022-1040.

Tussyadiah, L., Zach, F. (2017). Identifying salient attributes of peer-to-peer accommodation experience. *Journal of Travel & Tourism Marketing 34*(5), 636-652.

Varma, A., Jukic, N., Pestek, A., Shultz, C.J., Nestorov, S. (2016). Airbnb: Exciting innovation or passing fad? *Tourism Management Perspectives 20*, 228-237.

Vecchio, P.D., Lele, G., Ndou, V., Secundo, G. (2018). Creating value from Social Big Data: Implications for Smart Tourism Destinations. *Information Processing and Management 54*(5), 847-860.

von Helversen, B., Abramczuk, K., Kopeć, W., Nielek, R. (2018). Influence of consumer reviews on online purchasing decisions in older and younger adults. *Decision Support Systems 113*, 1-10.

Walsh, B. (2011). Today's Smart Choice: Don't Own. Share. http://content.time.com/time/specials/packages/article/0,28804,2059521_2059717,00.html (Accessed 13 August 2018).

Wang, D., Nicolau, J. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management 62*, 120-131.

Wu, C.-H., Ho, G.T.-S., Ho, D.C.-K, Chan, C.-Y.(2018). From Your Preferences to Niche Tourism: A New "To-Do" List in Hong Kong. In S. Sabah (Ed.), *Tourism*, IntechOpen, DOI: 10.5772/intechopen.82318.

Wu, C.H., Ho, G.T.S., Lam, C.H.Y., Ip, W.H., Choy, K.L., Tse, Y.K. (2016). An online niche-market tour identification system for the travel and tourism industry. *Internet Research 26*(1), 167-185.

Xie, K.L., Kwok, L. (2017). The effects of Airbnb's price positioning on hotel performance. *International Journal of Hospitality Management 67*, 174-184.

Zervas, G., Proserpio, D., Byers, J.W. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research 54*(5), 687-705.

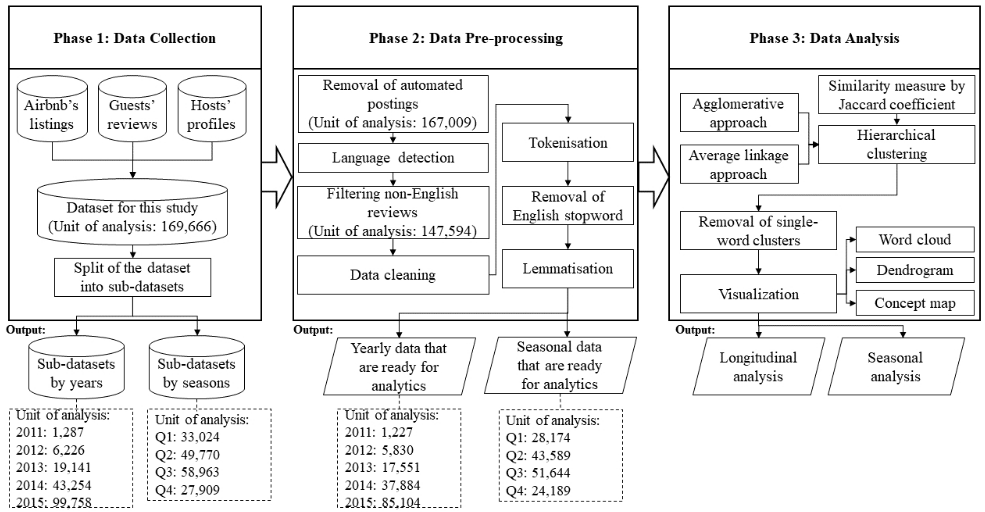

Phase 1: Data Collection
Airbnb's listings
Guests' reviews
Hosts' profiles
Dataset for this study (Unit of analysis: 169,666)
Split of the dataset into sub-datasets
Output:
Sub-datasets by years
Sub-datasets by seasons
Unit of analysis: 2011: 1,287 2012: 6,226 2013: 19,141 2014: 43,254 2015: 99,758
Unit of analysis: Q1: 33,024 Q2: 49,770 Q3: 58,963 Q4: 27,909
Phase 2: Data Pre-processing
Removal of automated postings (Unit of analysis: 167,009)
Language detection
Filtering non-English reviews (Unit of analysis: 147,594)
Data cleaning
Tokenisation
Removal of English stopword
Lemmatisation
Output:
Yearly data that are ready for analytics
Seasonal data that are ready for analytics
Unit of analysis: 2011: 1,227 2012: 5,830 2013: 17,551 2014: 37,884 2015: 85,104
Unit of analysis: Q1: 28,174 Q2: 43,589 Q3: 51,644 Q4: 24,189
Phase 3: Data Analysis
Agglomerative approach
Average linkage approach
Similarity measure by Jaccard coefficient
Hierarchical clustering
Removal of single-word clusters
Visualization
Word cloud
Dendrogram
Concept map
Output:
Longitudinal analysis
Seasonal analysis

Figure 1

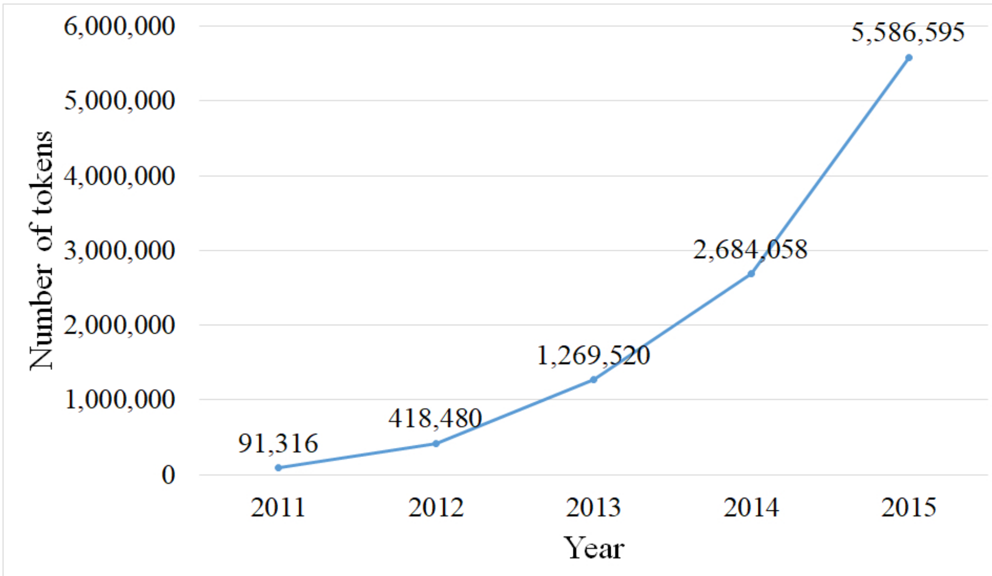

6,000,000
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
0
Number of tokens
91,316
418,480
1,269,520
2,684,058
5,586,595
2011
2012
2013
2014
2015
Year

Figure 2

Figure 3

165x128mm (72 x 72 DPI)

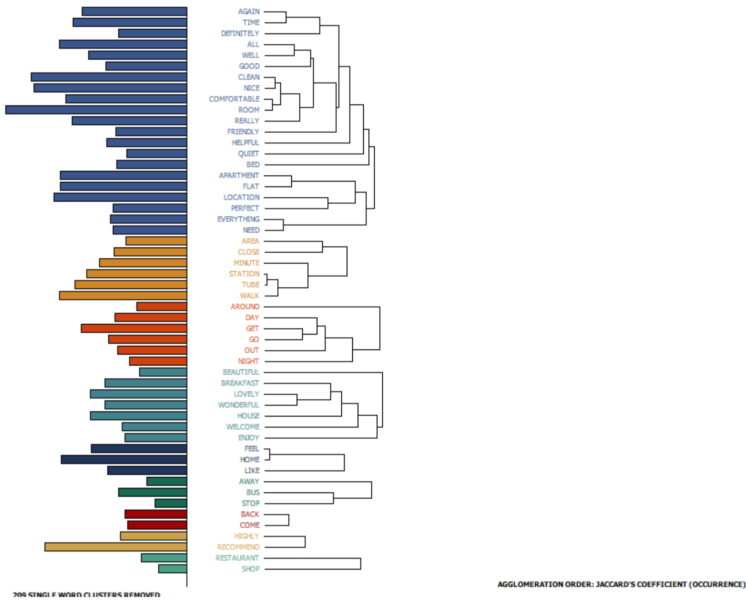
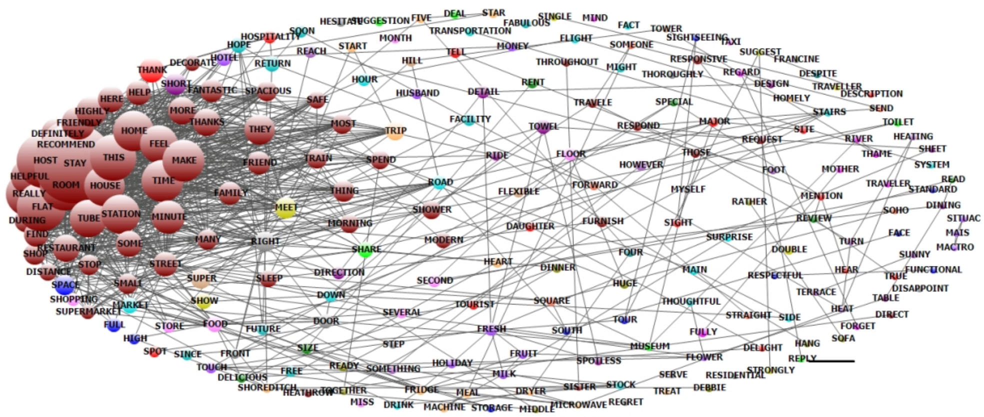


Figure 4

387x231mm (72 x 72 DPI)



HESITATE
SUGGESTION
FIVE
DEAL
STAR
SINGLE
MIND
FABULOUS
FACT
TOWER
SOON
TRANSPORTATION
SIGHTSEEING
TAXI
HOSPITALITY
MONTH
MONEY
FLIGHT
SOMEONE
SUGGEST
FRANCINE
HOPE
REACH
START
TELL
RESPONSIVE
HOTEL
HILL
THROUGHOUT
MIGHT
DESPITE
THANK
DECORATE
RETURN
THOROUGHLY
REGARD
TRAVELLER
HOUR
RENT
DESIGN
DESCRIPTION
SHORT
FANTASTIC
SPACIOUS
HUSBAND
DETAIL
HOMELY
HERE
HELP
SAFE
TRAVELE
SPECIAL
SEND
HIGHLY
MORE
STAIRS
FRIENDLY
THANKS
FACILITY
TOWEL
RESPOND
MAJOR
TOILET
DEFINITELY
HOME
THEY
MOST
TRIP
SITE
RECOMMEND
FEEL
REQUEST
RIVER
HEATING
HOST
STAY
THIS
MAKE
FRIEND
TRAIN
SPEND
RIDE
FLOOR
THOSE
THAME
SHEET
HOWEVER
SYSTEM
HELPFUL
ROOM
HOUSE
TIME
ROAD
FORWARD
FOOT
MOTHER
READ
REALLY
FAMILY
THING
FLEXIBLE
MYSELF
TRAVELER
STANDARD
FLAT
MEET
SHOWER
RATHER
MENTION
DINING
DURING
TUBE
STATION
MINUTE
MORNING
DAUGHTER
FURNISH
SIGHT
REVIEW
SOHO
SITUAC
FIND
MANY
RIGHT
MODERN
SURPRISE
FACE
MAIS
SHOP
RESTAURANT
SOME
SHARE
FOUR
DOUBLE
TURN
MACTRO
STOP
STREET
HEART
DINNER
MAIN
SUNNY
DISTANCE
DIRECTION
RESPECTFUL
HEAR
TRUE
FUNCTIONAL
SPACE
SMALL
SUPER
SLEEP
SECOND
HUGE
TERRACE
DISAPPOINT
SHOPPING
SHOW
DOWN
TOURIST
SQUARE
THOUGHTFUL
TABLE
SUPERMARKET
MARKET
SEVERAL
STRAIGHT
SIDE
HEAT
DIRECT
FULL
STORE
FOOD
FUTURE
DOOR
FRESH
SOUTH
TOUR
FULLY
FORGET
HIGH
STEP
MUSEUM
DELIGHT
HANG
SOFA
SPOT
SINCE
FRONT
SIZE
FRUIT
FLOWER
REPLY
TOUCH
FREE
READY
SOMETHING
HOLIDAY
SPOTLESS
STRONGLY
DELICIOUS
MILK
SERVE
RESIDENTIAL
SHOREDITCH
HEATHROW
TOGETHER
FRIDGE
MEAL
DRYER
SISTER
STOCK
TREAT
DEBBIE
MISS
DRINK
MACHINE
STORAGE
MIDDLE
MICROWAVE
REGRET

Figure 5

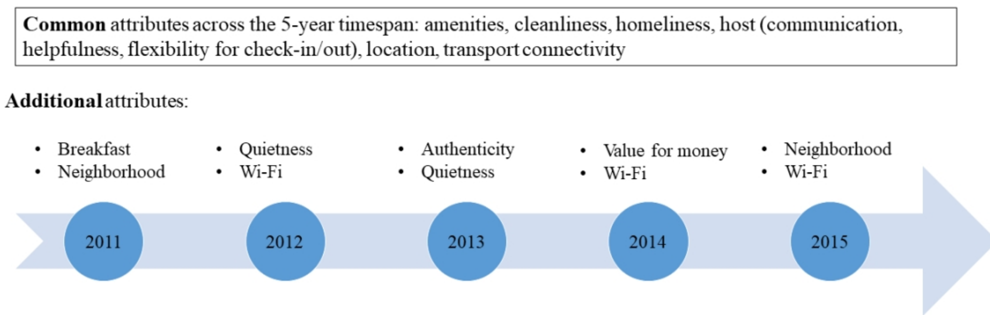

Common attributes across the 5-year timespan: amenities, cleanliness, homeliness, host (communication, helpfulness, flexibility for check-in/out), location, transport connectivity
Additional attributes:
• Breakfast
• Neighborhood
2011
• Quietness
• Wi-Fi
2012
• Authenticity
• Quietness
2013
• Value for money
• Wi-Fi
2014
• Neighborhood
• Wi-Fi
2015

Figure 6

List of Tables

Table 1. Description of the data set used in 7 similar papers

Reference	Sample size	Data type	Source(s)	Geographic scope	Longitudinal analysis?	Seasonal analysis?
Varma et al. (2016)	347 12	Primary data	Survey Interview	Not defined	×	×
Bae et al. (2017)	411	Primary data	Survey	South Korea	×	×
Brochado et al. (2017)	1,776	Secondary data	Online review	India, Portugal and the U.S.	×	×
Wang & Nicolau (2017)	298,331	Secondary data	Airbnb's listing	33 cities	×	×
Blah et al. (2018)	11	Secondary data	Airbnb's listing	San Francisco	×	×
Ju et al. (2018)	16,340 322	Secondary data Primary data	Online review Survey	Miami, New York, San Francisco and Chicago	×	×
Cheng & Jin (2019)	181,263	Secondary data	Online review	Sydney	×	×
This study	169,666	Secondary data	Online review	London	✓	✓

Table 2. Size of data collected for this study

Year	Q1	Q2	Q3	Q4	Total
2011	211	356	514	206	1,287
2012	1,347	1,549	2,405	925	6,226
2013	3,779	5,818	6,494	3,050	19,141
2014	8,425	12,770	14,888	7,171	43,254
2015	19,262	29,277	34,662	16,557	99,758
Total	33,024	49,770	58,963	27,909	169,666

Table 3. Size of data written in English only

Year	Q1	Q2	Q3	Q4	Total
2011	207	340	480	200	1,227
2012	1,273	1,454	2,244	859	5,830
2013	3,395	5,397	5,962	2,797	17,551
2014	7,391	11,246	12,950	6,297	37,884
2015	15,908	25,152	30,008	14,036	85,104
Total	28,174	43,589	51,644	24,189	147,596

Table 4. Frequency and TF-IDF of the top ten keywords

	Frequency	% Shown	% Processed	% Total	No. of Cases	% Cases	TF - IDF
LONDON	39175	2.05%	1.45%	0.70%	29625	34.81%	17953.6
FLAT	34394	1.80%	1.27%	0.62%	21550	25.32%	20516.1
PLACE	33473	1.75%	1.24%	0.60%	25026	29.41%	17792.9
ROOM	32827	1.72%	1.22%	0.59%	25830	30.35%	16998.7
HOST	32061	1.68%	1.19%	0.57%	29569	34.74%	14719.6
CLEAN	31205	1.63%	1.16%	0.56%	29342	34.48%	14431.1
APARTMENT	30110	1.58%	1.12%	0.54%	19344	22.73%	19372.9
NICE	29606	1.55%	1.10%	0.53%	23650	27.79%	16464.5
LOCATION	28814	1.51%	1.07%	0.52%	27231	32.00%	14259.7
RECOMMEND	24672	1.29%	0.91%	0.44%	24198	28.43%	13475.1

Table 5. Recommendations for hosts in different seasons

Season	Recommendations for hosts
Q1	Highlight if the listing has an in-house washing machine (that allows guests to use) or any laundry services nearby so that guests can keep their clothes clean; provide clear instructions on how to use the washing machine.
Q2	The listing is more attractive to guests if guests can enjoy sunbathing or swimming nearby.
Q3	*[Q3 does not have particular discussion themes]*
Q4	Highlight the amenities such as hot water that can give guests a warm and cosy stay; provide fast internet connection and TV so that guests can entertain when indoor.